MOTION IMAGERY STANDARDS BOARD **TECHNICAL REFERENCE MATERIAL** **Constructing a MISP Compliant File/Stream**	**MISB TRM 0909.4** **23 October 2014**

1 Scope

An example MISP (Motion Imagery Standards Profile) compliant Motion Imagery (MI) file/stream constructed from technologies approved by the Motion Imagery Standards Board is presented as an aid in understanding. This document reflects the revised MISP 2015, which states requirements for systems to meet MISP compliance. This document is informative only. If there are any discrepancies between the two, the MISP takes precedence.

2 References

[1] MISP-2015.1 Motion Imagery Standards Profile, Oct 2014
[2] MISB RP 0904.1 H.264 Bandwidth/Quality/Latency Tradeoffs, Feb 2014
[3] MISB RP 0605.4 Time Stamping and Metadata Transport in High Definition Uncompressed Motion Imagery, Feb 2014
[4] MISB RP 0603.2 Common Time Reference for Digital Motion Imagery Using Coordinated Universal Time (UTC), Feb 2014
[5] MISB ST 0604.3 Time Stamping Compressed Motion Imagery, Feb 2014
[6] MISB RP 0802.2 H.264/AVC Motion Imagery Coding, Feb 2014
[7] MISB ST 0807.14 MISB KLV Metadata Dictionary, Oct 2014
[8] MISB ST 0607.3 MISB Metadata Registry and Processes, Oct 2014
[9] MISB TRM 1006 Key-Length-Value (KLV) Users Guide, Jun 2012
[10] MISB ST 0902.4 Motion Imagery Sensor Minimum Metadata Set, Oct 2014
[11] MISB ST 0601.8 UAS Datalink Local Set, Oct 2014
[12] MISB ST 0102.11 Security Metadata Universal and Local Sets for Digital Motion Imagery, Oct 2014
[13] MISB ST 1402 MPEG-2 Transport of Compressed Motion Imagery and Metadata, Feb 2014
[14] MISB ST 0804.4 Real Time Protocol for Motion Imagery and Metadata, Feb 2014

3 Acronyms

AAF	Advanced Authoring Format
AVC	Advanced Video Codec (H264)
DTS	Decoding Time Stamp
EG	Engineering Guideline
ES	Elementary Stream

Kb	kilo-bits
kB	kilo-bytes
KLV	Key-Length-Value
MI	Motion Imagery
MISP	Motion Imagery Standards Profile
MXF	Material Exchange Format
PED	Processing-Exploitation-Dissemination
PES	Packetized Elementary Stream
PID	Packet IDentifier
PTS	Presentation Time Stamp
RP	Recommended Practice
RTP	Real Time Protocol
RTSP	Real Time Streaming Protocol
SDI	Serial Data Interface
SMPTE	Society for Motion Picture and Television Engineers
ST	Standard
TRM	Technical Reference Material
TS	MPEG-2 Transport Stream
UTC	Coordinated Universal Time
VANC	Vertical Ancillary Data
VBI	Vertical Blanking Interval

4 Revision History

Revision	Date	Summary of Changes
0904.4	10/23/2014	Updated for consistency to MISP 2015.1

5 Introduction

MISP 2015.1 [1] states:

> "Compliance is achieved when the requirements – expressed using statements with the word "shall" are met. Numerous standards are referenced within these requirements; such references contain requirements themselves, which are to be followed to be MISP compliant. Standards build on other standards, and so compliance to a standard must satisfy all the references indicated within their subsequent requirements."

A MISP compliant file/stream for Motion Imagery has the following three components with timing/synchronization implicit in the Metadata and/or Container:

Motion imagery: compressed (the visual content)

Metadata: Key-Length-Value form (e.g. mission, security, optional data)

Media Container: approved package that carries MI, metadata, or both. In a MPEG-2 transport stream (TS) both MI and metadata must be present. In Real-Time Transport Protocol (RTP), MI and metadata are independently carried as separate streams with appropriate timing that associate the individual media streams.

A "file" is generally considered a complete media package, such as a MI clip stored on a server. A "stream" could be a live event transmitted directly from a sensor, or a pre-recorded file that is subsequently played out in continuous fashion. Both files and streams are composed of Motion Imagery, metadata and optionally audio that are containerized into: a multiplexed MPEG-2 Transport Stream; media-specific RTP streams, or a MXF/AAF file.

6 File/Stream Construction - example

As way of illustration, the building blocks comprising a MISP compliant file/stream are shown in Figure 1. This depiction may not indicate how these functions are implemented, or what in what order the operations may occur.

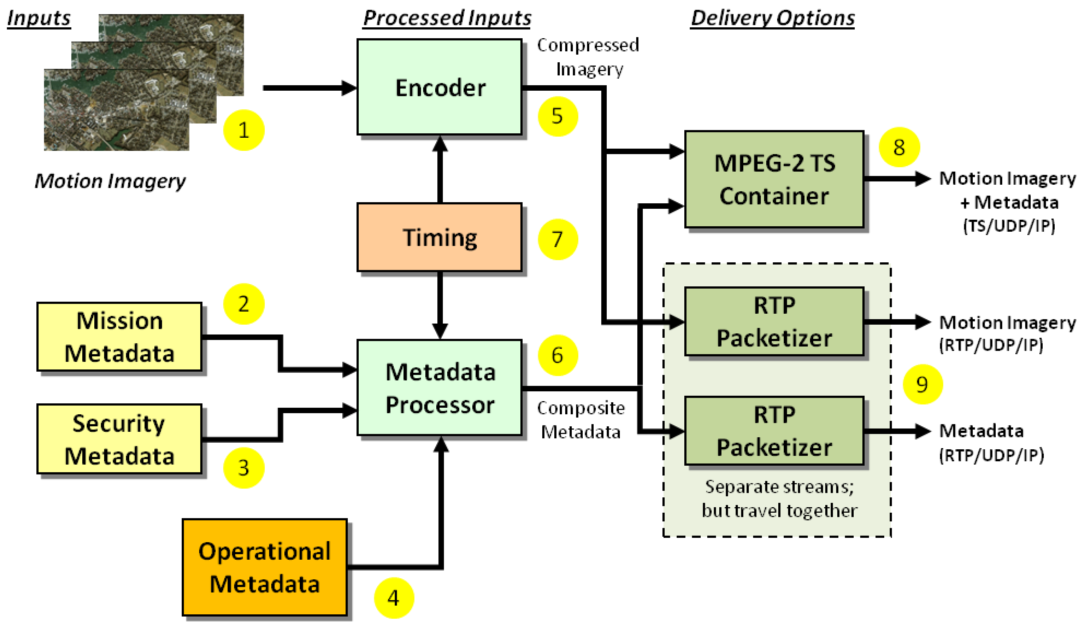


Figure 1: A General Structure for Constructing a MISP-Compliant File/Stream

1 Motion Imagery is produced by an Imager (typically non-compressed)

2 Mission metadata (geo position, sensor angles, etc.) is generated by a mission computer

3 Security metadata (required) is supplied

4 Optional metadata (e.g. chat, audio, annotation) may be added

5 An encoder compresses the Motion Imagery (e.g. H.264, MPEG-2)

6 All metadata is multiplexed together in preparation for its encapsulation into a Container

7 Absolute Time Stamps are inserted into both the MI and metadata

8 MI and metadata may be combined into a MPEG-2 Transport Stream (TS) Container

9 Optionally, MI and metadata may be disseminated as independent RTP streams

Table 1 identifies the functional elements needed to create a MISP compliant Motion Imagery stream/file. The dashed lines (Transport Vehicle) indicate that these data may be carried either together or separately. The MISP [1] is very specific in defining methods for data transport, allowed compression types, metadata encodings, semantics and elements, timing standards and relationships, and data packaging using standardized containers. These details are more thoroughly defined within the MISP and associated documents.

Table 1: Media Elements, Timing, and Container

Source Data Container	Media Components	Relationship	Media Container
Transport Vehicle	Compressed Motion Imagery	Time Stamping Metadata composition	Transport Vehicle
Transport Vehicle	Metadata		Transport Vehicle

MISP Compliant File Streams

Table 2 shows that Motion Imagery may be supplied over a SDI (Serial Digital Interface) – for example, the *Transport Vehicle* to an H.264/AVC encoder *Media Component*. The Metadata *Media Component* may be supplied within the same SDI signal, or via some other input protocol means, such as Ethernet, or RS-232. The relationship between the Motion Imagery and Metadata media components is maintained through timing assigned and contained within each media component. The two media components may then be packaged together – as in a MPEG-2 TS, or as individual single components – as in RTP, depending on the desired *Transport Vehicle*.

Table 2: Specific Technologies for Electro-optical and Infrared Imagery

<table>
<tr><th>Source Data</th><th colspan="2">Media Components</th><th>Relationship</th><th>Output Data</th></tr>
<tr><th>Transport</th><th colspan="2">Compressed MI</th><th>Timing/Multiplex</th><th>Transport</th></tr>
<tr><td rowspan="5">HD-SDI
SDI
Other</td><td colspan="2" rowspan="2">H.264/AVC
MPEG-2</td><td rowspan="5">MPEG-2 TS
RTP</td><td>File</td></tr>
<tr><td>MPEG-2 TS
AAF/MXF</td></tr>
<tr><td colspan="2">Metadata</td><td rowspan="2">Stream</td></tr>
<tr><td>Mission</td><td>Security</td></tr>
<tr><td colspan="2">KLV</td><td>MPEG-2 TS
RTP</td></tr>
</table>

7 Example Implementation: *A Digital HD MI Sensor System*

Figure 2 shows an example High Definition digital Motion Imagery collection system aligning with the processing steps in Figure 1. In Appendix A, an example legacy analog sensor system is shown. Various MISB standards that are relevant to each step are indicated for reference.

Step 1: MI Source – sensor and channel

The format of the High Definition (HD) imager drives many system design considerations: this includes the spatial dimensions (e.g. 1280x720 or 1920x1080 pixels), the temporal update rate (i.e. 30, 60 or other images per second), the pixel value range (e.g. 8, 10 bits per pixel), and the color sampling model (i.e. 4:4:4, 4:2:2, etc.). **MISB RP 0904 - H.264 Bandwidth / Quality / Latency Tradeoffs** [2] offers suggested methods to meet system bandwidth constraints by reducing the spatial image size (scaling), reducing the image field of view (cropping), or reducing the temporal rate.

These criteria guide the choice of transport for moving the content from the sensor to the next stage of processing, for example, compression. SMPTE provides several standards for the transport of HD non-compressed video, all based on serial data protocols; these are known as HD-SDI (Serial Data Interface) and 3G-SDI, both for HD.

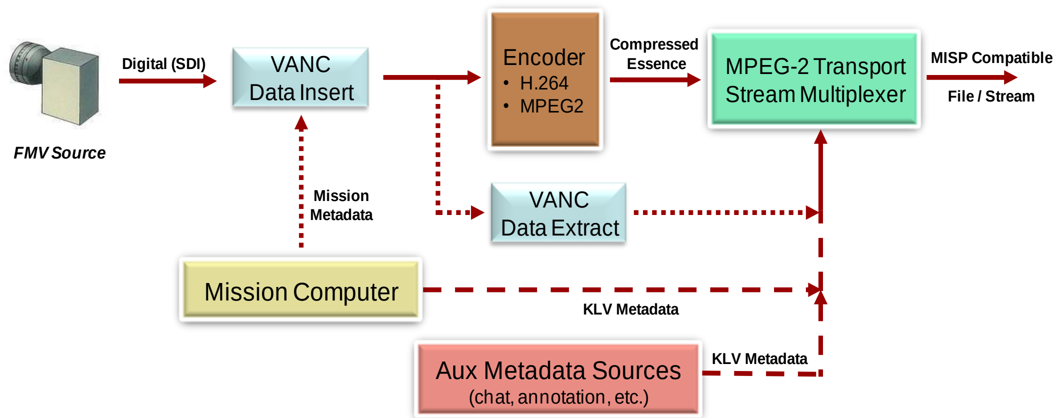


Figure 2: Example Digital Motion Imagery System

Step 2: *VANC Data Insert – sensor interface*

Within the appropriate SDI Container, space is available for inserting non-imagery data. This data space is called the ancillary interval, and the MISP mandates that metadata when placed into this space specifically use the Vertical Ancillary data space or VANC (see Figure 3). **MISB ST 0605 - Time Stamping and Metadata Transport in High Definition Uncompressed Motion Imagery** [3] defines how to map metadata into the VANC. The VANC can be populated on an image frames basis; this allows metadata to be added to each frame, if desired.

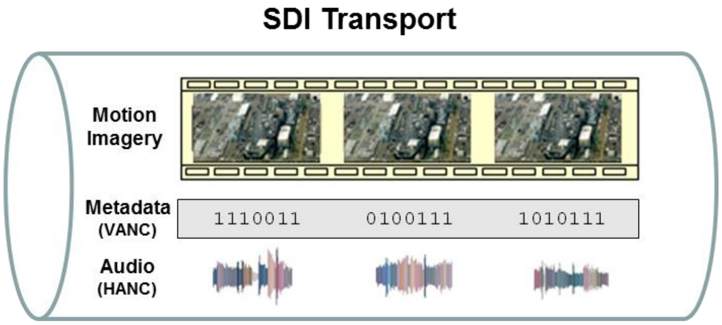


Figure 3: SDI Transport for up to three types of data

Figure 3 shows the SDI interface, which provides carriage for several types of media: Motion Imagery, metadata, and audio. Note that the metadata is carried within the VANC; audio is carried within the HANC (Horizontal Ancillary data space). Not all metadata (if any) may emanate from the sensor subsystem, but some newer sensor systems embed metadata into the VANC within the sensor architecture. In this example, mission metadata is inserted into the VANC for transport to the encoder.

Within the VANC Metadata a Precision Time Stamp is inserted as metadata; this is described in MISB ST 0605. The Precision Time Stamp is based on UTC (Coordinated Universal Time) and is defined in **MISB ST 0603 - Common Time Reference for Digital Motion Imagery Using Coordinated Universal Time (UTC)** [4]. The Precision Time Stamp is also used in populating the time stamp metadata element contained in many MISB metadata sets.

Step 3: *Encoder – Motion Imagery compression*

The High Definition Motion Imagery is compressed using one of the allowed compression types, which includes H.264/AVC and MPEG-2. MISP 2015.1 provides Points of Interoperability (POI) for H.264/AVC systems and MPEG-2 Legacy systems. The POI specify spatial and temporal formats that assure maximum interoperability within the community.

In the compression process, the Precision Time Stamp is extracted from the VANC or mission computer, and inserted into a defined data space within the compressed stream. **MISB ST 0604 - Time Stamping Compressed Motion Imagery** [5] provides guidance on the format and placement of the Precision Time Stamp. **MISB RP 0802 - H.264/AVC Motion Imagery Coding [6]** provides parameters and application-specific modes for H.264/AVC deemed more optimal within our community of practice.

Step 4: *Metadata – mission, security, and auxiliary*

Metadata inserted into the VANC is extracted and input to a MPEG-2 transport stream multiplexer. Metadata directly from a mission computer is likewise input to the MPEG-2 transport stream multiplexer.

All metadata must conform to the MISB standards for syntax and semantics. SMPTE KLV (Key-Length-Value) is the encoding for MISB encoded on an air platform. This binary format was chosen because of its efficiency, extendibility and decodable robustness. The MISB maintains a listing of all metadata keys in **MISB ST 0807 - KLV Metadata Dictionary** [7] with a corresponding standard **MISB ST 0607 - MISB Metadata Registry and Processes** [8] that describes the fundamentals of the metadata dictionary. A tutorial on KLV structures and rules can be found in **MISB TRM 1006 - Key-Length-Value (KLV) Users Guide** [9].

While many MISB KLV elements describe important mission data, not all are mandated to meet be MISP compliant. Certain mission metadata and security metadata must, however, be present.

Mission Metadata

MISB ST 0902 - Motion Imagery Sensor Minimum Metadata Set [10] specifies the mandated set of KLV elements that characterize many of the dynamic parameters collected during a mission. This set is drawn from a more complete set defined in **MISB ST 0601 - UAS Datalink Local Set** [11]. The 0902 minimum set enables basic discovery and retrieval functionality in exploitation. While certain mission metadata may be updated at the Motion Imagery image rate, "Metadata elements contained in the Minimum Metadata Set **shall** be reported no less than once every thirty seconds (30) under all circumstances" (see ST 0902).

Security Metadata

The security Metadata set is defined in **MISB ST 0102 - Security Metadata Universal and Local Sets for Digital Motion Imagery** [12] and must be present within the metadata in a file or stream to be MISP compliant. It is recommended that this metadata be placed early in the file/stream to help determine the security level of the data without searching through the entire dataset, and that its repetition within the file/stream occur at least every 10 seconds (ST 0902). *A point of clarity: the required elements of the local security set as defined in ST 0102 are identified within ST 0902. Thus, choosing to define security using ST 0902 satisfies the requirement for MISP compliance.*

Auxiliary Metadata (optional)

Other types of metadata may be inserted at the sensor, or anywhere within the processing/exploitation workflow. Numerous MISB standards for metadata such as chat, annotation, and audio are documented in various MISB ST and RP's. Attention to the additional overhead in bandwidth these types of metadata may incur is necessary; more non-imagery data means less data for Motion Imagery, which could impact image quality. Refer to the MISB website for more information on these various metadata constructs.

Step 5: *MPEG-2 Transport Stream Multiplexer – timing and transport*

Once the High Definition MI in compressed form (compressed essence) and the KLV metadata populated, the two need to be packaged into a Container for delivery (or storage). In the example of Figure 1, the two media types are combined into a MPEG-2 Transport Stream. **MISB ST 1402 - MPEG-2 Transport of Compressed Motion Imagery and Metadata** [13] specifies requirements on the use of the TS.

Synchronous/Non-synchronous Metadata Carriage

Two methods for metadata carriage are described in ST 1402; a synchronous method and a non-synchronous method. Among other differences, the synchronous method allows for a

Presentation Time Stamp (PTS) to be applied to the metadata. This method provides a means to synchronize events in the Motion Imagery with metadata for presentation simultaneously at a display, for example. Although the PTS of the Motion Imagery and that of the Metadata can be made the same, it is also possible to force the PTS values to align with the absolute time stamps, if desired.

It is important to note that the PTS values are not intended to provide a timing relationship between the Motion Imagery and the metadata for exploitation and analysis. A PTS is a *relative* timing mechanism intended to align various media streams together for presentation purposes only.

8 The MPEG-2 Transport Stream – an overview

Step A – *the elementary stream*

To understand how data is carried in a transport stream it is instructive to start at the beginning of the process with non-compressed source imagery.

Figure 4 shows a series of non-compressed digital Motion Imagery images that get compressed.

Motion Imagery conforming to MPEG standards is composed of a series of full, predicted, and bi-directionally predicted frames referred to as I, P, and B frames (image) respectively. I (Intraframe) frames (or full pictures) are coded based on information within a frame independent of surrounding frames and, therefore, generally compress the least (80kB). P (Predictive) frames are predicted from I frames (or other P frames), and thus, compress more than I frames (20kB). Finally, B (Bi-directional) frames are predicted from both I and P frames in a bi-directional fashion using image information from frames both before and after. B frames generally compress the most (5kB); however, their dependence on neighboring frames makes them very susceptible to the propagation of errors in reconstruction. I, P, and B compressed frames are called Access Units. Step A shows a coding structure of one I, two B, and one P frame. This pattern, while not unique, will repeat over the entire coded sequence. Note also that the compressed data values are for illustration only and will differ substantially as a function of image scene complexity and encoder algorithm.

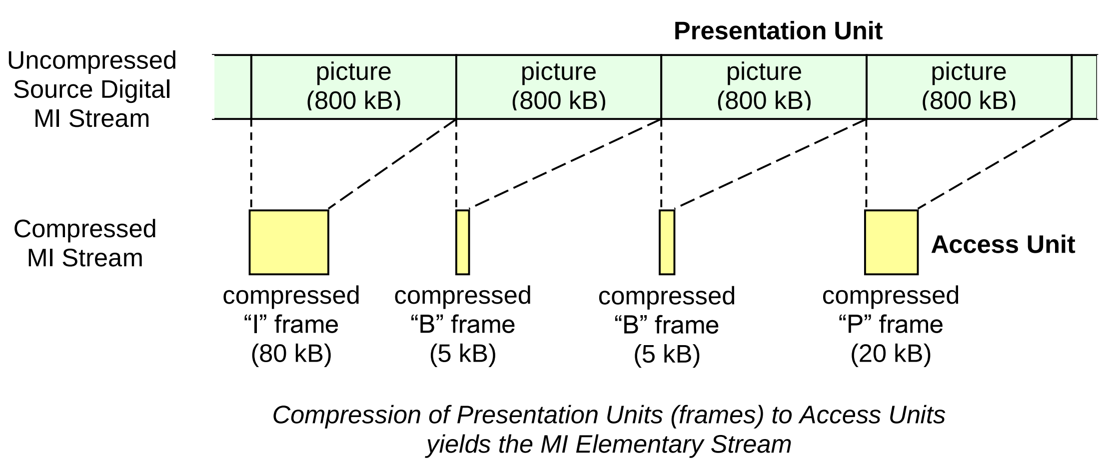


Figure 4: Step A - the compression process

Step B – *creating a packetized elementary stream*

Once the source non-compressed imagery is compressed into Access Units it must be prepared for packaging into the MPEG-2 transport stream. Step B in the process (Figure 5) is to create packetized elementary stream (PES) packets of each compressed stream (remember there may be multiple streams of Motion Imagery, audio, metadata, etc.) with accompanying information that indicates what is contained in the packet and how it should be decoded. A PES packet header carries this control information.

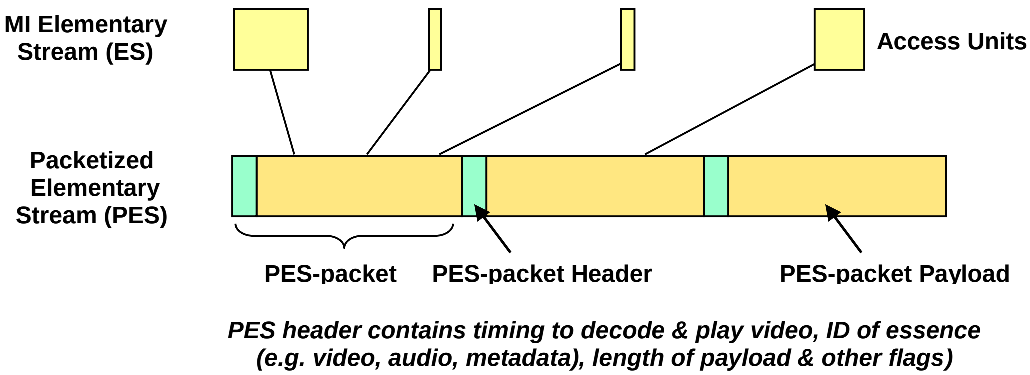


Figure 5: Step B - packetizing the elementary stream

PES packets can vary in length and contain a number of Access Units. Within the PES header a decode time stamp (DTS) aids in decoding the data. A presentation time stamp (PTS) aids in the presentation of the data to the display. The PTS is especially important in synchronizing the different media types (Motion Imagery, audio, and metadata) at the display. Appendix B provides a list of various flags and fields in a PES packet.

Step C – *mapping into transport stream packets*

PES packets are mapped into the transport stream itself. Each TS packet is fixed in size (188 bytes). A four-byte TS header contains various indicators, such as a start code for the TS packet, an error indicator for bad packets, a packet ID or PID that identifies what media the packet of data represents (see Figure 6). This information is required so that a decoder can decipher the data appropriately. Appendix B lists various flags and fields for the TS header.

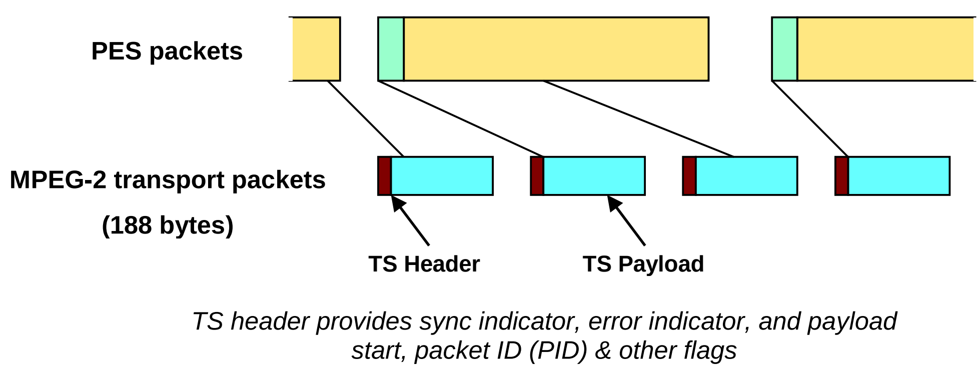


Figure 6: Step C - mapping PES packets to transport stream packets

Finally, in Step D (Figure 7) the transport stream multiplexer formats the individual packetized media streams into transport-stream-format packets, and interleaves them into one composite stream of data. The data rate of the final TS stream will reflect the aggregate sum of respective individual media components. The 'Null' packet serves to fill in gaps in the stream to help maintain a desired transport stream data rate.

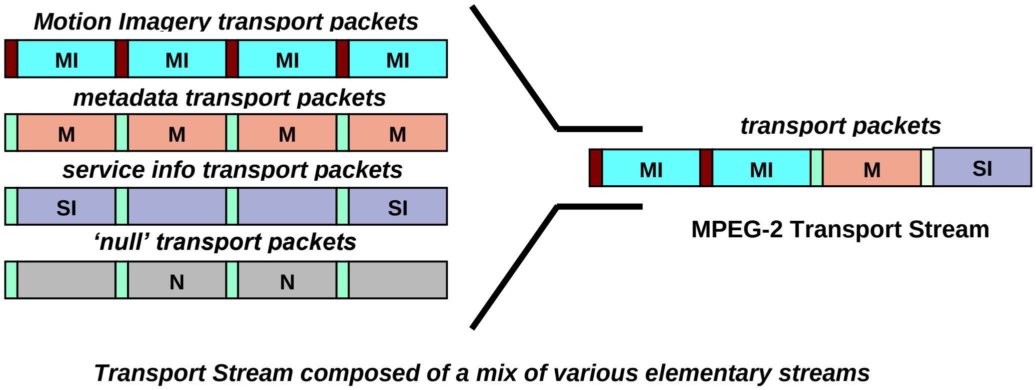
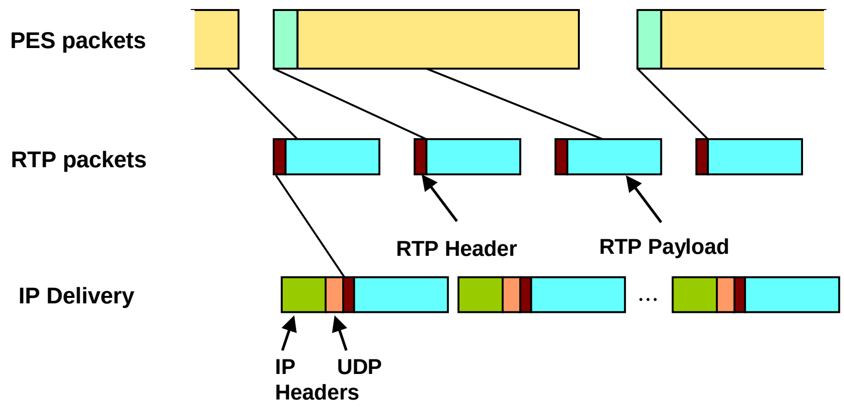


Figure 7: Step D - multiplexing media transport packets into one stream

Although the process above illustrates how Motion Imagery is mapped into the transport stream, the metadata is similarly packetized and receives a PES header. Once packets of Metadata are constructed they are then mapped into transport stream packets with a corresponding PTS if the streams are intended to be displayed together with tight timing constraints.

9 Real Time Transport Protocol

The packaging of media for RTP follows the same first two steps of the process above for transport stream. Figure 8 shows an additional step where the packetized data (PES packets) receive a RTP header on each PES packet, which then is encapsulated into an IP datagram with IP and UDP headers. For more information on RTP and a prototype implementation see **MISB ST 0804 - Real-Time Protocol for Motion Imagery and Metadata** [14].



IP, UDP, RTP headers precede data composed of PES packets

Figure 8: Packaging one media component for RTP delivery

10 File/Stream Protocol

The Motion Imagery and the metadata together represent the collected "asset". To be of value this asset must be kept intact within either a unified Container, or with appropriate timing, to facilitate downstream analysis. There are several methods available within the MISB suite of technologies: MPEG-2 TS (Transport Stream), RTSP/RTP, and AAF/MXF. These different methods are useful in different applications. MPEG-2 TS is the dominant Container in the MISB community. It serves as both a Container for point-to-point transport and also for storage.

11 Advice on Getting Started

The MISB web site (http://www.gwg.nga.mil/misb/index.html) publishes all the documents needed to construct a MISP compliant file or stream. The MISP is the authoritative MISB reference. It contains references to all the Standards (ST), Recommended Practices (RP), Engineering Guidelines (EG), and Technical Reference Material (TRM) supported by the MISB, and also provides up-to-date industry references foundational to the MISB documents.

While the MISB provides many standards, not all are necessary for a given application. Of most importance is diligence in following the particular metadata standard – particularly the required Minimum Metadata Set (ST 0902) and Security Set (ST 0102). TRM 1006 and ST 0601 are instructive in understanding how KLV packets are encoded in KLV, and other topics like platform data flow and error detection. ST 0604 addresses time stamping and time code for the Motion Imagery, and time stamping of metadata. ST 1402 addresses synchronization of the imagery with metadata, and how to carry metadata along with Motion Imagery in a MPEG-2 TS.

A Motion Imagery asset is much more valuable when it is accompanied with metadata; collecting pure MI alone is of little use downstream for exploitation. <u>MISP compliant streams must contain metadata</u>. Mission metadata is critical for assisting discovery and retrieval of an asset, and for relating one asset or intelligence data to another. The MISB provides encompassing metadata sets to support mission metadata. How to add enriched metadata, such as annotation, chat, audio, etc. is also provided through various MISB standards. As more Motion Imagery is manipulated throughout the network with processes—such as transcoding for bandwidth-challenged edge delivery—the MISB will continue to look to industry for best practices and adopt or leverage those found applicable.

It is advisable to revisit the MISB web site periodically for updates and additions that may impact implementation decisions. The MISB encourages community participation and feedback in shaping the recommendations to best fit community needs. The goal is interoperability with standards as foundational and clarity an objective.

The MISB supplies test files that represent various compression and metadata combinations for both standard and high definition Motion Imagery. These First Level Integration (FLI) files are generally less than one minute in duration. The metadata is typically random with respect to the actual Motion Imagery content. Although random, the metadata exercises all the keys and tags of a particular MISB metadata set with data values that are permitted. FLI test files are available on the protected MISB site for community use in testing products and systems.

Questions or misinterpretations of MISB documentation should be brought to the attention of the MISB. Periodic meetings (three per year) are held to discuss such issues and provide appropriate changes where needed.

12 Appendix A – An Analog NTSC Sensor (Informative)

While use of analog in any new deployments is strongly discouraged this section is intended to provide continuity in understanding of the processes from legacy system to digital system implementations.

Figure A1 illustrates a simplified sensor collection system where analog video is captured along with mission metadata. In this example, it is assumed that the encoding, metadata processing, and container packaging is done at the ground station.

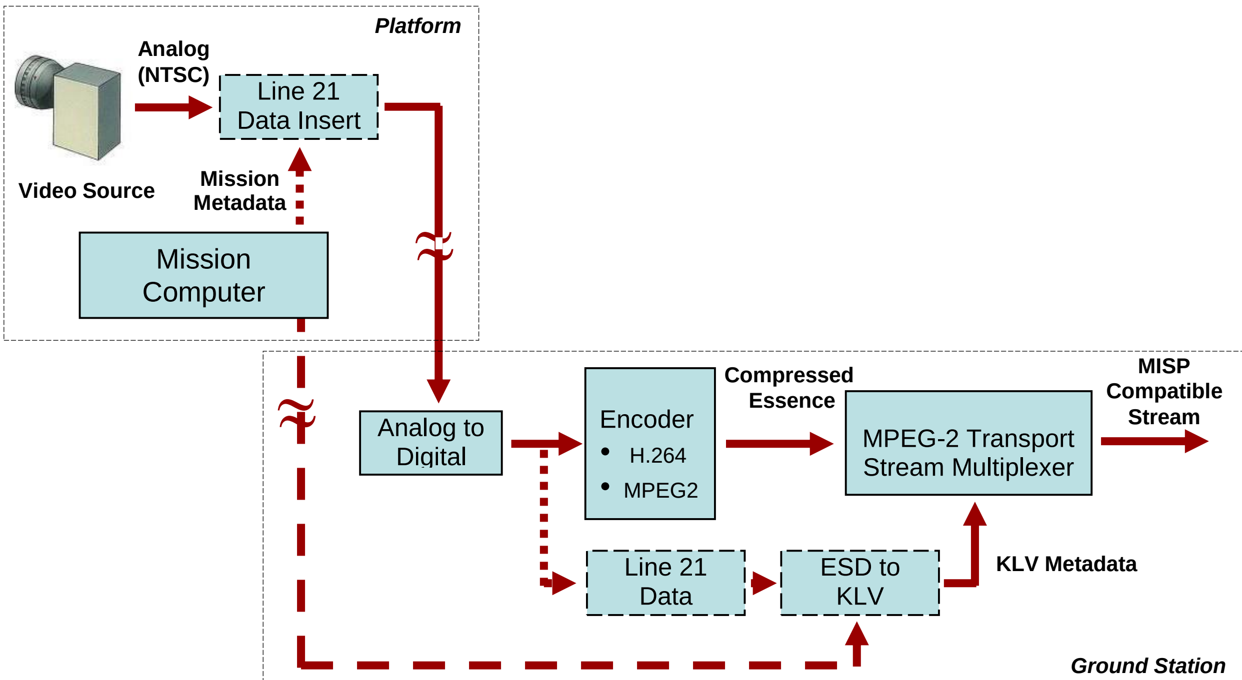


Figure A1: Analog Video Sensor System

Mission metadata may have been inserted into the analog video on Line 21 (designated for television closed captioning), or transmitted directly to the ground station through an independent path. Once at the ground station, the analog video is converted to digital and then encoded using H.264/AVC or MPEG-2 compression. Metadata (in ESD format) inserted on Line 21 at the platform will be extracted and converted to KLV. Metadata that arrives via an independent path from the platform will likewise be converted to KLV. In both cases the KLV metadata will be multiplexed with the compressed Motion Imagery into a MPEG-2 Transport Stream container for dissemination or storage.

Let's examine some of the elements in this process further. The NTSC analog television standard supports 525 lines for one complete video frame. This is delivered to the display as two successive "fields," where each field is 262.5 lines, and reconstructed as complete frames in a process called interlace. Think of a numbered list of lines that make up an image where in one field the odd lines are "painted" followed by a second field that paints the even lines. Because the display phosphors have persistence and retain a portion of the previous field the illusion to the viewer is a flicker-free rendition of the picture.

The NTSC standard allows for roughly 482 lines to be used for picture content with the remaining lines—called the Vertical Blanking Interval (VBI)—hidden behind the bezel of the display. Line 21 is one such unused picture line. With the introduction of closed captioning Line 21 became one of the designated places to carry this data. This is also where metadata is often placed in analog capture systems. Mission metadata is sometimes modulated onto Line 21 as a means to carry the data with the video.

At the ground station the analog video is converted to digital through an analog-to-digital converter. This produces a number of samples called pixels (picture elements) horizontally. The vertical direction is already “digital” in that each scan line is discrete and independent of adjacent ones. Each pixel is typically quantized to eight bits of gray-scale resolution and the color components may be quantized to eight but typically fewer bits, such as four.

The digitized video signal is ready for compression through the encoder. In parallel, Line 21 of the analog video is stripped of any metadata that may have been inserted at the platform and converted to KLV format. Any metadata that comes directly from the platform independent of the video is also converted, if necessary. All KLV metadata is finally input to the transport stream multiplexer along with the compressed video essence.

The function of the transport stream multiplexer is to insert the two (or more) different data sources into one common package. One can think of the transport stream as containing a series of parallel tracks where each track can carry its own data type—analogous to a multi-lane highway. The “multiplex” serves to merge the various data lanes into one—think of merging traffic at an on-ramp. Certain track identifiers and management directories are placed into the transport stream to guide the extraction of the various data types at the decoder.

Similarly to the digital MI example in the text optional metadata may be added to the composite multiplexed MI file/stream anywhere along the PED workflow. Such metadata may include chat, audio, annotations and other data types. See the MISB website http://www.gwg.nga.mil/misb/index.html for documents that guide such metadata additions.

13 Appendix B – TS, PES Headers (Informative)

Tables B1 and B2 list various flags and data fields that comprise the transport stream (TS) header and the packetized elementary stream (PES) header (ISO/IEC 13818-1 [26] Tables 2-2, 2-21) respectively. Required values for the carriage of metadata as indicated in **ST 1402 – MPEG-2 Transport of Compressed Motion Imagery and Metadata** are shown in the metadata-only tables (Tables B3, B4), as are values that although optional are recommended for consistency in usage.

Table B1: Transport Stream Header ***(4 bytes minimal)***			
	No. of bits	***shall***	***recommended***
sync_byte	8	0x47	
transport_error_indicator	1		usage based
payload_unit_start_indicator	1		usage based
transport_priority	1		'0'
PID	13		usage based
transport_scrambling_control	2	'00' (not scrambled)	
adaptation_field_control	2		usage based
additional flags, data possible	K bytes		usage based
continuity_counter	4		usage based

Table B2: Packetized Elementary Stream Header			
	No. of bits	***shall***	***recommended***
packet_start_code_prefix	24	0x000001	
stream_id	8		usage based
PES_packet_length	16		usage based
'10'	2	'10'	
PES_scrambling_control	2	'00' (not scrambled)	
PES_priority	1		'0' (low)
data_alignment_indicator	1		usage based
copyright	1		'0' (not specified)
original_or_copy	1		'1' (original)
PTS_DTS_flags	2		usage based
ESCR_flag	1		usage based
ES_flag_rate	1		usage based
DSM_trick_mode_flag	1	'0'	No trick mode field
additional_copy_info_flag	1		usage based

Table B2: Packetized Elementary Stream Header			
PES_CRC_flag	1		usage based
PES_extension_flag	1		usage based
PES_header_data_length	8		usage based
If (PTS_DTS_flags == '10')			
'0010'	4	'0010'	fixed
PTS[32..30]	3		usage based
marker_bit	1	'1'	
PTS[29..15]	15		usage based
marker_bit	1	'1'	
PTS[14..0]	15		usage based
marker_bit	1	'1'	fixed
If (PTS_DTS_flags == '11')			
'0010'	4	'0010'	fixed
PTS[32..30]	3		usage based
marker_bit	1	'1'	fixed
PTS[29..15]	15		usage based
marker_bit	1	'1'	fixed
PTS[14..0]	15		usage based
marker_bit	1	'1'	fixed
'0001'	4	'0001'	fixed
DTS[32..30]	3		usage based
marker_bit	1	'1'	fixed
DTS[29..15]	15		usage based
marker_bit	1	'1'	fixed
DTS[14..0]	15		usage based
marker_bit	1	'1'	fixed
If ESCR_flag == '1'			
Reserved	2	'11'	reserved
ESCR_base[32..30]	3		usage based
marker_bit	1	'1'	fixed
ESCR_base[29..15]	15		usage based
marker_bit	1	'1'	fixed
ESCR_base[14..0]	15		usage based
marker_bit	1	'1'	fixed
ESCR_extension	9		usage based
marker_bit	1	'1'	fixed
If (ES_rate_flag == '1')			
marker_bit	1	'1'	fixed
ES_rate	22		usage based
marker_bit	1	'1'	fixed
If (additional_copy_info_flag == '1')			

Table B2: Packetized Elementary Stream Header			
marker_bit	1	‘1’	fixed
additional_copy_info	7		usage based
If (PES_CRC_flag == ‘1’)			
previous_PES_packet_CRC	16		usage based
If (PES_extension_flag == ‘1’)			
PES_private_data_flag	1		usage based
pack_header_field_flag	1		usage based
program_packet_sequence_counter_flag	1		usage based
P-STD_buffer_flag	1		usage based
Reserved	3	‘111’	reserved
PES_extension_flag_2	1		usage based
If (PES_private_data_flag == ‘1’)			
PES_private_data	128		usage based
If (pack_header_field_flag == ‘1’)			
pack_field_length	8		usage based
If (program_packet_sequence_counter_flag == ‘1’)			
marker_bit	1	‘1’	fixed
program_packet_sequence_counter	7		usage based
marker_bit	1	‘1’	fixed
MPEG1_MPEG2_identifier	1		usage based
original_stuff_length	6		usage based
If (P-STD_buffer_flag == ‘1’)			
‘01’	2	‘01’	fixed
P-STD_buffer_scale	1		usage based
P-STD_buffer_size	13		usage based
If (PES_extension_flag_2 == ‘1’)			
marker_bit	1	‘1’	fixed
PES_extension_field_length	7		usage based
stream_id_extension_flag	1		usage based
If (stream_id_extension_flag == ‘0’)			
stream_id_extension	7		usage based
data	N x 8		usage based
stuffing_byte	N1 x 8		usage based
PES_packet_data_byte	N2 x 8		usage based

Table B3: PES Header for Asynchronous Metadata Carriage			
	No. of bits	***shall***	***recommended***
packet_start_code_prefix	24	0x000001	

stream_id	8	0xBD (private stream 1)	
PES_packet_length	16		usage based
‘10’	2	‘10’	
PES_scrambling_control	2	‘00’ (not scrambled)	
PES_priority	1		‘0’ (low)
data_alignment_indicator	1		usage based
copyright	1		‘0’ (not specified)
original_or_copy	1		‘1’ (original)
PTS_DTS_flags	2	‘00’ (no PTS/DTS)	
ESCR_flag	1	‘0’ (not present)	
ES_flag_rate	1	‘0’ (not present)	
DSM_trick_mode_flag	1	‘0’ (not present)	
additional_copy_info_flag	1	‘0’ (not present)	
PES_CRC_flag	1	‘0’ (not present)	
PES_extension_flag	1	‘0’ (not present)	
PES_header_data_length	8	N1 (bytes)	
stuffing_byte	N1 x 8		N1 bytes
PES_packet_data_byte	N2 x 8		N2 bytes

Table B4: PES Header for Synchronous Metadata Carriage

	No. of bits	shall	recommended
packet_start_code_prefix	24	0x000001	
stream_id	8	0xFC (metadata stream)	
PES_packet_length	16		usage based
‘10’	2	‘10’	fixed
PES_scrambling_control	2	‘00’ (not scrambled)	
PES_priority	1		‘0’ (low)
data_alignment_indicator	1		usage based
copyright	1		‘0’ (not specified)
original_or_copy	1		‘1’ (original)
PTS_DTS_flags	2	‘10’ (PTS only)	
ESCR_flag	1	‘0’ (not present)	
ES_flag_rate	1	‘0’ (not present)	
DSM_trick_mode_flag	1	‘0’ (not present)	
additional_copy_info_flag	1	‘0’ (not present)	
PES_CRC_flag	1	‘0’ (not present)	
PES_extension_flag	1	‘0’ (not present)	

Table B4: PES Header for Synchronous Metadata Carriage			
PES_header_data_length	8	5+N1 (bytes)	usage based
If (PTS_DTS_flags == ‘10’)		‘10’ (PTS only)	
‘0010’	4	‘0010’	fixed
PTS[32..30]	3		usage based
marker_bit	1	1	fixed
PTS[29..15]	15		usage based
marker_bit	1	1	fixed
PTS[14..0]	15		usage based
marker_bit	1	1	fixed
stuffing_byte	N1 x 8		N1 bytes
PES_packet_data_byte	N2 x 8		N2 bytes